DEPTH-BASED FACE RECOGNITION USING LOCAL QUANTIZED PATTERNS ADAPTED FOR RANGE DATA

Tomás Mantecón Carlos R. del-Blanco Fernando Jaureguizar Narciso García

ABSTRACT

A depth-based face recognition algorithm specially adapted to high range resolution data acquired by the new Microsoft Kinect 2 sensor is presented. A novel descriptor called Depth Local Quantized Pattern descriptor has been designed to make use of the extended range resolution of the new sensor. This descriptor is a substantial modification of the popular Local Binary Pattern algorithm. One of the main contributions is the introduction of a quantification step, increasing its capacity to distinguish different depth patterns. The proposed descriptor has been used to train and test a Support Vector Machine classifier, which has proven to be able to accurately recognize different people faces from a wide range of poses. In addition, a new depth-based face database acquired by the new Kinect 2 sensor have been created and made public to evaluate the proposed face recognition system.

***Index Terms*—** Kinect 2, Depth Local Quantized Pattern, face recognition, face database, classification.

1. INTRODUCTION

Face recognition systems have many practical applications, such as people awareness in controlled environments, entrance allowance in security area, identification of missing/searched people, or human-computer interaction. For this reason, a wide range of works have been published in the area of computer vision with the aim of recognizing or identifying people faces using different feature extraction techniques and classification solutions [1]. Most of the existing approaches use color imagery to achieve the recognition goal [2]. Others complement the color information with complex and costly 3D-face models to be more robust [3]. There also exist some works that combine color and depth imagery to increase the accuracy in the recognition [4]. In these last cases, the depth information is just a complement of the color information since the existing depth imagery has such a low range/depth resolution that prevents to use it independently or in substitution of the color imagery.

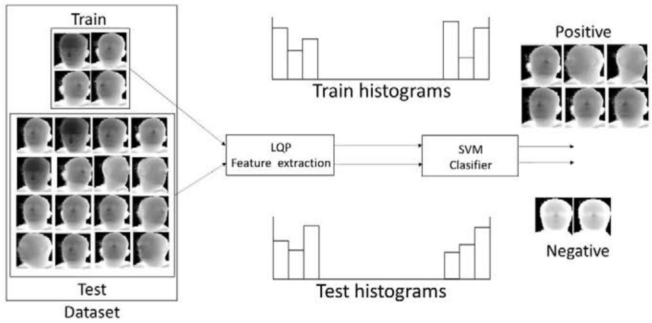

Fig. 1. DLQP-SVM algorithm.

Regarding the state-of-the-art techniques that use color imagery, popular feature extraction techniques to characterize people faces are: the Scale Invariant Feature Transform (SIFT) [5], the Histogram of Oriented Gradients (HOG) [6], and the Local Binary Pattern (LBP) [7]. These image feature descriptors are used in combination with different classification techniques for the purpose of face recognition, such as Support Vector Machines (SVM) [7], AdaBoost [6], or Neural Networks [8]. The approaches that also make use of 3D information usually use the same descriptors and classification techniques. For example, LBP and SIFT descriptors are used in [3] along with an SVM classifier, although other simpler classification approaches can be found such as the Nearest-Neighbor algorithm [9].

On the other hand, although a lot of work has been done in the last years with the Microsoft Kinect 1 [10], a low-cost depth camera, it has serious limitations for depth based face recognition. The high level of noise and the reduced resolution in depth are the main problems since make almost impossible to extract reliable features from the depth relief of eyes, nose, mouth, ears, etc. However, a second generation of depth sensors is arriving with extended depth/range resolution that can have promising applications for face recognition based on depth information. Senz3D [11] from Creative and Kinect 2 [12] from Microsoft (the second version of Kinect sensor) formed part of this new generation.

In this paper, a novel face recognition algorithm that only

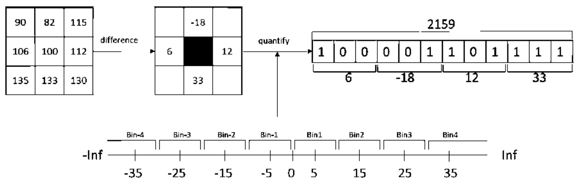

Fig. 2. DLQP descriptor.

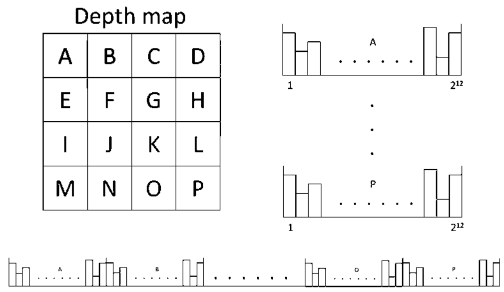

Fig. 3. DLQP-based feature descriptor computation. See the text for more details.

uses the depth information acquired from the new Kinect 2 is presented. The paper presents two major contributions. The first one is the design of a novel descriptor called Depth Local Quantized Pattern descriptor (DLQP) that makes use of the extended range resolution of the Kinect 2. This descriptor is a significant variation of the popular Local Binary Pattern algorithm, widely used in face recognition with color imagery. The key is the introduction of a quantification step that exploits the higher range resolution of the Kinect 2 data, which increases the capacity to distinguish different depth patterns. The proposed DLQP has been used along with an SVM classifier for the recognition of people faces that have a wide range of poses. Fig. 1 outlines the proposed recognition algorithm. The second contribution is the creation of a public depth-based face database acquired with the Kinect 2 sensor that is used to validate the performance of the proposed algorithm and other state-of-the-art approaches. As far as the authors' knowledge, it is the first database that has this kind of information.

The organization of the paper is as follows. Section 2 describes the Depth Local Quantized Pattern descriptor that is used to characterize the faces. Section 3 presents the SVM classification technique that uses as input features the DLQP descriptors. A description of the database is presented in Section 4. The obtained results with the proposed algorithm are presented in Section 5. Finally, conclusions are drawn in Section 6.

2. DLQP BASED FEATURE EXTRACTION

The LBP [13] is a widely used algorithm for feature extraction in many face recognition systems. The main idea of this technique is to calculate the difference between the intensity value of the central pixel and the intensity value of some pixels in its neighborhood. Then, the computed differences are encoded by a binary value. The main advantages are the invariance to illumination changes and the relatively low computational cost. Some extensions have been proposed to be robust to rotations and changes in scale. It has been also used with depth or 3D information, but without any special adaptation.

Inspired on LBP, a new descriptor, called DLQP, specifically adapted to high depth resolution information has been designed. The main modification is that the difference between the depth value of the central pixel and the neighborhood ones is quantified, making richer the characterization of the depth information. Consequently, this fact increases the capacity of distinguishing depth patterns because depth values are not so affected by illumination changes as color imagery. The quantification process has been adapted to the purpose of face characterization and recognition, limiting the range of depth differences to quantify. The limitation of the quantification fulfills other functional restriction: keep the DLQP-based feature vector enough short to be used in practice. According with these design lines, only the differences between -35 mm and 35 mm have been quantized, since they represent the range of the most relevant depth differences in people faces. Other design criterion has been to use a uniform quantification with 8 intervals, i.e. 3 bits are needed to encode every quantized depth difference. In addition, only 4 neighbors (west, north, east, and south) are considered to keep the DLQP descriptors with a reasonable size.

More formally, let it be (x_c, y_c) the coordinates of a pixel in the image, the computation of a DLQP code can be expressed as follows:

$$DLQP_{P,R}(x_c, y_c) = \text{stack}_{P=0}^{P-1} q(d_p - d_c), \tag{1}$$

where R is the radius of the neighborhood that has P surrounding pixels, d_c and d_p are the depth values of the central pixel and one of the P pixels of the neighborhood, respectively. The function $q(d_p - d_c)$ computes the quantification of the difference between the depth values d_p and d_c, and finally stack is a function that simply stack the binary code resulting from the difference quantification.

Every $DLQP_{P,R}(x_c, y_c)$ is a number that contributes to a bin of a histogram with $N_b \times P$, where N_b is the number of bits used in the quantification process. In this case, $N_b = 3$ and $P = 4$, resulting in a histogram of 2^{12} values (bins). Figure 2 shows the quantification and bin computation process.

To increase the relevance of the spatial information, which is lost with the histogram approach, every depth-based face image is divided into 16 blocks of equal size, and then a different DLQP histogram is computed for each one. Finally, the DLQP-based feature descriptor is obtained by concatenating all the histograms in a single vector. Figure 3 summarizes the process.

3. CLASSIFICATION PROCESS

To implement the classification process, the Pegasos SVM technique [14] has been used. A configuration of one-vs-all have been adopted to solve an identification problem (i.e. identify to which subject belongs each face). To improve the results of Pegasos solution, a Hellinger kernel, more commonly known as Bhattacharyya distance [15], has been used

$$k(h, h') = \sum_{i} \sqrt{h(i)h'(i)}, \qquad (2)$$

where h and h' are the DLQP-based feature descriptors (described in the previous section) of the test and training data faces, respectively.

To perform the classification task, the classifier has been trained with positive and negative DLQP-based feature descriptors. To make this possible, the used database have been divided into a training set and a test one, using a 20% of the depth images for the training process and the other 80% for the testing process.

4. DEPTH-BASED FACE DATASET

As far as the authors' knowledge, there is no a face database based on high depth resolution data, i.e. with a resolution significantly higher than the one obtained by the first Kinect sensor [16]. In consequence, a new database acquired by the Kinect 2 sensor have been acquired. This database is composed by 22 sequences of 18 different subjects (15 men and 3 women, 4 people with glasses). They were recorded while each subject was sitting about 50 cm away from the sensor and the head was at the same height as the sensor. All subjects rotated their heads with the aim of obtaining as much information as possible from the head: nose, eyes, mouth, ears, etc. The database has been published at the web https://sites.google.com/site/hrrfaced/, and it is publicly available under the name High Resolution Range based Face Database (HRRFaceD).

Figure 4 shows three sets of depth images from the HRRFaceD database, where the main characteristics of the face can be distinguished: eyes, mouth, ears and nose. Depth maps have been saved with 16 bits of depth resolution, and with a spatial resolution of 512×424 pixels.

5. RESULTS

The proposed face recognition algorithm, from now on called DLQP-SVM, has been tested and compared with other state-

(a)
(b)
(c)

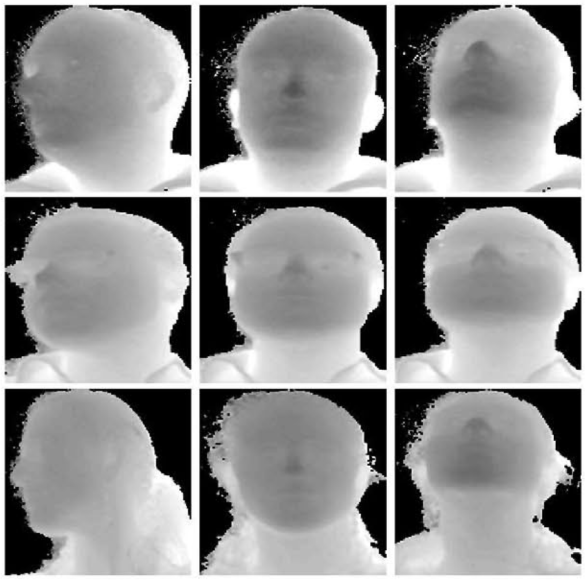

Fig. 4. Three sets of depth maps for different subjects from the HRRFaceD database, acquired by the Kinect 2 sensor: (a) a man without glasses, (b) a man with glasses and (c) a woman without glasses.

of-the-art algorithms using the HRRFaceD database presented in the previous section. A part of depth maps obtained by the Kinect 2 sensor of size 180×180 pixels has been selected, as can be show in Figure 4. It is important to notice that different face poses of each subject have been used, which increases the difficulty of the recognition process. To extract the features of each image, they have been split into 16 blocks of 45×45 pixels (see Figure 3) according to the explanation in Section 2. There are 200 frames for each subject, which have been split into 40 train frames, 160 test frames.

The results of the DLQP-SVM algorithm has been compared with an LBP-based method proposed in [7], and a SIFT-based method proposed in [5]. The LBP method uses the LBP without any modification. The SIFT one is based on the use of dense SIFT feature extraction. To be able to compare the results between different methods, the confusion matrix (CM) is used. The CM is widely used in object recognition to measure the recognition performance. Each column represents the number of predicted faces belonging to each class, and each row represents the total number of outcomes (positives and negatives) that belongs to each class. Each element of the main diagonal of that matrix represents the accuracy of each class, which is expressed as follows:

$$\text{Accuracy} = 100 \times \frac{\text{Total number of correct faces}}{\text{Total number of faces}}, \qquad (3)$$

With the CM it is also possible to obtain the false discovery rate (fdr) (i.e. the number of false positives) for each face. That measure is the sum of the elements of each column of

Sub.	01	02	03	04	04*	05	05*	06	07	08	09	09*	10	11	12	13	14	15	16	17	17*	18
01	67,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	13,13	0,00	0,00	0,00	3,75	15,63	0,00	0,00	0,00	0,00	0,00
02	0,00	76,88	0,00	0,00	0,00	0,00	0,00	0,00	6,88	0,00	0,00	0,00	0,00	10,63	0,00	0,00	0,00	0,00	5,63	0,00	0,00	0,00
03	0,00	0,00	66,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	10,63	7,50	0,00	15,63	0,00	0,00	0,00	0,00	0,00	0,00	0,00
04	0,00	0,00	0,00	57,50	10,63	0,00	0,00	0,00	5,63	0,00	0,00	0,00	8,75	0,00	14,38	0,00	0,00	0,00	0,00	3,13	0,00	0,00
04*	0,00	0,00	0,00	0,00	73,75	0,00	0,00	0,00	13,13	0,00	0,00	0,00	0,00	0,00	13,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00
05	0,00	0,00	0,00	0,00	0,00	34,38	30,00	0,00	0,00	0,00	3,13	3,13	0,00	0,00	16,88	0,00	0,00	12,50	0,00	0,00	0,00	0,00
05*	0,00	0,00	0,00	0,00	0,00	0,00	77,50	0,00	0,00	0,00	0,00	22,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
06	0,00	0,00	3,13	0,00	4,38	0,00	3,13	57,50	0,00	0,00	0,00	4,38	6,88	2,50	10,00	0,00	0,00	0,00	3,75	0,00	4,38	0,00
07	0,00	0,00	0,00	0,00	4,38	0,00	0,00	0,00	50,63	0,00	0,00	0,00	0,00	11,88	0,00	0,00	15,63	0,00	17,50	0,00	0,00	0,00
08	0,00	0,00	0,00	0,00	3,13	0,00	15,00	0,00	0,00	68,75	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	13,13	0,00	0,00
09	5,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	6,88	0,00	62,50	18,75	3,75	0,00	3,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00
09*	0,00	0,00	0,00	0,00	0,00	0,00	5,00	5,63	4,38	0,00	4,38	66,88	0,00	0,00	10,63	0,00	0,00	0,00	3,13	0,00	0,00	0,00
10	26,25	0,00	4,38	0,00	0,00	0,00	3,13	0,00	5,63	0,00	0,00	4,38	51,88	0,00	0,00	0,00	4,38	0,00	0,00	0,00	0,00	0,00
11	0,00	0,00	3,13	0,00	0,00	0,00	4,38	0,00	3,75	0,00	0,00	0,00	0,00	49,38	27,50	0,00	3,13	5,00	3,75	0,00	0,00	0,00
12	0,00	0,00	0,00	13,75	0,00	0,00	3,13	0,00	0,00	0,00	0,00	3,75	0,00	5,00	38,75	0,00	28,13	0,00	3,13	4,38	0,00	0,00
13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	6,25	8,13	81,88	3,75	0,00	0,00	0,00	0,00	0,00
14	5,63	0,00	0,00	3,13	0,00	0,00	0,00	0,00	3,13	4,38	0,00	0,00	3,75	3,13	13,13	10,63	53,13	0,00	0,00	0,00	0,00	0,00
15	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	9,38	0,00	0,00	0,00	71,88	0,00	18,75	0,00	0,00
16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,38	0,00	0,00	0,00	87,50	0,00	8,13	0,00
17	0,00	0,00	0,00	0,00	0,00	0,00	3,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00	63,13	9,38	19,38
17*	0,00	0,00	0,00	0,00	0,00	0,00	3,75	0,00	0,00	0,00	0,00	15,00	0,00	4,38	0,00	0,00	0,00	0,00	0,00	7,50	69,38	0,00
18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	20,00	0,00	80,00

Table 1. Confusion matrix obtained using the DLQP-SVM algorithm.

the CM except the one that belongs to the main diagonal and normalize it by the sum of all the elements of the column, this measure is expressed as follows:

$$\text{fdr} = 100 \times \frac{\text{Total number of false positive faces}}{\text{Total number of faces}}, \tag{4}$$

For the comparison with the LBP algorithm, the following parameters for the DLQP-SVM algorithm have been used: radius of the neighbor $R = 1$, number of neighbors $P = 4$ (west, north, east and south), and 3 bits for the quantifier (only for the DLQP-SVM). With these parameters the length of the final DLQP-SVM descriptor is a histogram of $16 \times 2^{4\times3} = 65536$ bins. For the SIFT algorithm, the parameters proposed in [5] have been used.

The results of the proposed DLQP-SVM algorithm using the HRRFaceD database are shown in Table 1. Each row and each column represent each class (face) of the database. The case where the number of the subject has an '*' is because that subject is wearing glasses. As it can be observed, most of the accuracy values are greater than 60%. To evaluate those results it is important to consider that the pose of the subject's face is unknown and it is continuously changing. It also can be noticed that in those cases that we have the same subject but with and without glasses the most misclassification error goes to that subject.

Results of the proposed algorithm and the other state-of-the-art algorithms are shown in Table 2, in all cases using the proposed database. Values of the main diagonal of the confusion matrix obtained for each algorithm are shown. It can be noticed that our algorithm achieves better results than the LBP one, and slightly better than the SIFT one. The false discovery rate values indicates that the misclassification error is distributed along the other classes, i.e. there is no class that have a greater misclassification value.

Sub.	DLQP-SVM		LBP		SIFT	
	ac.	fdr	ac.	fdr	ac.	fdr
01	67,50	35,33	40,63	0,00	71,25	7,32
02	76,88	0,00	50,00	0,00	91,25	33,33
03	66,25	13,82	90,63	46,49	56,25	18,18
04	57,50	22,69	51,88	35,16	59,38	29,63
04*	73,75	23,38	40,00	0,00	75,63	41,26
05	34,38	0,00	50,00	41,61	37,50	47,83
05*	77,50	47,68	100,0(	66,87	54,38	53,23
06	57,50	8,91	56,25	0,00	45,00	41,94
07	50,63	49,38	40,00	0,00	66,88	14,40
08	68,75	5,98	52,50	0,00	83,75	17,79
09	62,50	10,71	47,50	0,00	48,13	44,60
09*	66,88	58,85	44,38	1,39	60,00	54,72
10	51,88	37,12	48,75	1,27	76,25	50,81
11	49,38	51,83	41,25	8,33	73,75	34,81
12	38,75	77,94	36,88	53,91	35,63	51,69
13	81,88	14,94	86,88	78,08	93,13	26,96
14	53,13	57,07	50,63	55,74	60,00	41,10
15	71,88	19,58	83,75	9,46	65,00	28,77
16	87,50	32,37	88,13	51,38	66,88	23,57
17	63,13	51,44	91,25	28,08	63,13	35,67
17*	69,38	23,97	53,75	10,42	65,00	45,83
18	80,00	19,50	45,63	0,00	56,25	21,05
mean	63,95	30,11	58,67	22,19	63,84	34,75

Table 2. Accuracy (ac.) and false discovery rate (fdr) results using the proposed algorithm (DLQP-SVM) and other state-of-the-art algorithms in the HRRFaceD database.

6. CONCLUSIONS

A face recognition algorithm using only depth information has been presented in this paper. The depth information has been obtained by a new generation of depth sensors: Kinect 2. The proposed algorithm extracts robust and high distinguishable features for the task of face recognition, achieving good classification results along with an SVM classifier. In addition, a new database with high depth resolution faces have been created to validate the performance of the proposed system.

7. REFERENCES

[1] Di Huang, Caifeng Shan, M. Ardabilian, Yunhong Wang, and Liming Chen, "Local binary patterns and its application to facial image analysis: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 6, pp. 765–781, 2011.

[2] Tao Gao, X.L. Feng, He Lu, and J.H. Zhai, "A novel face feature descriptor using adaptively weighted extended LBP pyramid," *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 23, pp. 6286 – 6291, 2013.

[3] Hengliang Tang, Baocai Yin, Yanfeng Sun, and Yongli Hu, "3d face recognition using local binary patterns," *Signal Processing*, vol. 93, no. 8, pp. 2190 – 2198, 2013.

[4] M. Pamplona, S. Sarkar, D. Goldgof, L. Silva, and O. Bellon, "Continuous 3d face authentication using rgb-d cameras," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013, pp. 64–69.

[5] S. Anith, D. Vaithiyanathan, and R. Seshasayanan, "Face recognition system based on feature extration," in *International Conference on Information Communication and Embedded Systems (ICICES)*, Feb. 2013, pp. 660–664.

[6] B. Jun, I. Choi, and D. Kim, "Local transform features and hybridization for accurate face and human detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1423–1436, Jun. 2013.

[7] G. Kayim, C. Sari, and C.B. Akgul, "Facial feature selection for gender recognition based on random decision forests," in *Signal Processing and Communications Applications Conference (SIU), 2013 21st*, Apr. 2013, pp. 1–4.

[8] O. Nikisins and M. Greitans, "Local binary patterns and neural network based technique for robust face detection and localization," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–6.

[9] Yiding Wang, Meng Meng, and Qingkai Zhen, "Learning encoded facial curvature information for 3d facial emotion recognition," in *Seventh International Conference on Image and Graphics (ICIG)*, Jul. 2013, pp. 529–532.

[10] "Microsoft Corporation. Kinect for Xbox 360 http://dx.doi.org/10.1007/s10107-010-0420-4," .

[11] "Creative Corporation. Creative Senz3D http://creative.com/p/web-cameras/creative-senz3d," .

[12] "Microsoft Corporation. Kinect for Xbox ONE http://www.microsoft.com/en-us/kinectforwindows/," .

[13] T. Ojala, M. Pietikinen, and M. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.

[14] S. Shalev-Shwartz, Y. Singer, N. Srebro, and A. Cotter, "Pegasos: primal estimated sub-gradient solver for svm," *Mathematical Programming*, vol. 127, no. 1, pp. 3–30, 2011.

[15] Euisun Choi and Chulhee Lee, "Feature extraction based on the bhattacharyya distance," *Pattern Recognition*, vol. 36, no. 8, pp. 1703 – 1709, 2003.

[16] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, "Random forests for real time 3d face analysis," *International Journal of Computer Vision*, vol. 101, no. 3, pp. 437–458, Feb. 2013.